L6 ANSWER 24 OF 34 BIOSIS COPYRIGHT (c) 2004 The Thomson Corporation. on STN
AN 1993:428737 BIOSIS
DN PREV199396083362
TI Engineering proteins for nonnatural environments.
AU Arnold, Frances H.
CS Div. Chem. Chemical Eng., Calif. Inst. Technol., Pasadena, CA 91125, USA
SO FASEB (Federation of American Societies for Experimental Biology) Journal, (1993) Vol. 7, No. 9, pp. 744-749.
CODEN: FAJOEC. ISSN: 0892-6638.
DT Article
LA English
ED Entered STN: 22 Sep 1993
Last Updated on STN: 22 Sep 1993
AB The ability to use proteins in nonnatural environments greatly expands their potential applications in biotechnology. Because nature has not paid much attention to optimizing proteins for in vitro applications under conditions that differ substantially from their natural surroundings, there is generally room for improvement through alterations in the amino acid sequence. The most effective approach to this protein engineering task depends on the level to which the molecular basis for the desired property is understood. Consistently successful "rational" design using site-directed mutagenesis requires a high level of understanding of structure and mechanisms or, alternatively, a particularly simple strategy for obtaining the desired feature. An example of a generally applicable and easy-to-implement protein stabilization strategy is metal ion chelation by specific surface dihistidine sites, which can affect thermal stability as well as the protein's ability to withstand denaturants such as guanidinium chloride. Random mutagenesis, on the other hand, can be effective even when structure or mechanisms are poorly understood, provided one can conveniently screen or select for the property of interest. This approach is illustrated by the sequential accumulation of random **mutations** that greatly enhance the catalytic activity of a **serine protease**, subtilisin E, in polar organic solvents. The random mutagenesis approach, which mimics the natural evolutionary refinement process, can be used to "coax" enzymes into tolerating nonnatural environments.

L11 ANSWER 48 OF 49 MEDLINE on STN DUPLICATE 22
AN 82280203 MEDLINE
DN PubMed ID: 7051707
TI [Thermitase, a thermostable **serine protease** of Thermoactinomyces vulgaris: interaction of the active center and the SH-group of the enzyme].
Thermitase, eine thermostabile Serin-Protease aus Thermo- actinomyces vulgaris: Wechselwirkung zwischen aktivem Zentrum und SH-Gruppe des Enzyms.
AU Hansen G; Frommel C; Hausdorf G; Bauer S
SO Acta biologica et medica Germanica, (1982) 41 (2-3) 137-44.
Journal code: 0370276. ISSN: 0001-5318.
CY GERMANY, EAST: German Democratic Republic
DT Journal; Article; (JOURNAL ARTICLE)
LA German
FS Priority Journals
EM 198210
ED Entered STN: 19900317
Last Updated on STN: 20000303
Entered Medline: 19821021
AB Modification of the serine and histidine residue in the active centre of thermitase with diisopropylfluorophosphate (DFP) or L-1-tosylamide-2-phenylethyl chloromethylketon (TPCK), and of the only SH-group of the enzyme, with Hg-compounds causes an activity loss against hydrolysis of 4-nitrophenylacetate. While the **modification** of **cysteine** prevents reaction of **serine** and histidine in the active centre of the enzyme with DFP and TPCK, respectively, the Hg2+- and CF3Hg+-binding to the SH-group after modification of essential amino acid residues in the active centre is retained. To elucidate the interaction of the SH-group with the active centre, the modified products of thermitase were investigated for their thermostability. Ca2+-ions were found to have a stabilizing effect on all the modified products of thermitase, as well as on the native enzyme. Simultaneous **modification** of the **cysteine** and **serine** leads to an increase in thermostability of thermitase, whilst double **modification** at the cysteine and histidine causes destabilization of the enzyme.

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

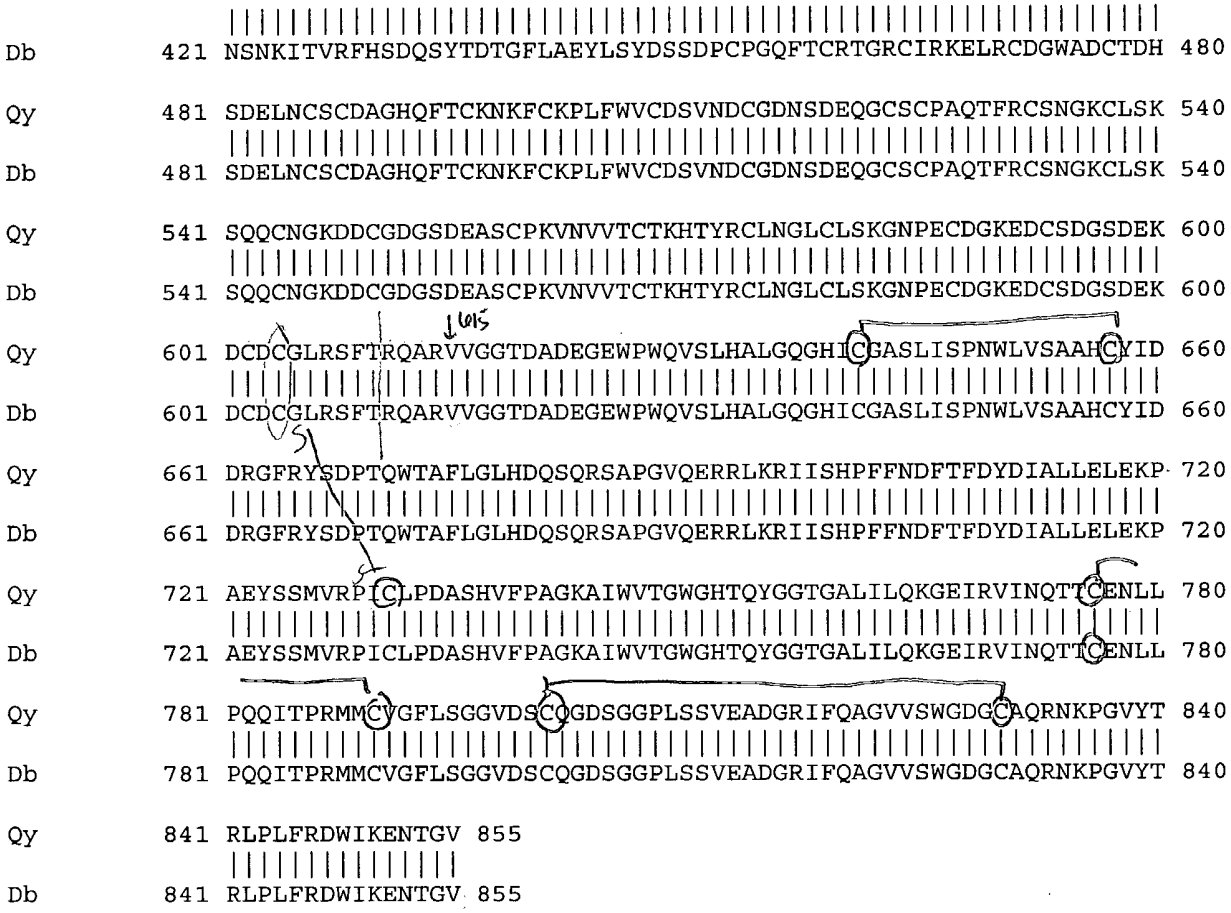
SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4681	100.0	855	2	US-09-027-337-2 5972616	Sequence 2, Appli
2	4681	100.0	855	4	US-09-644-600-2 6451500	Sequence 2, Appli
3	4681	100.0	855	4	US-09-654-600A-2 6649741	Sequence 2, Appli
4	3810	81.4	902	4	US-09-644-600-10	Sequence 10, Appl
5	3810	81.4	902	4	US-09-654-600A-10	Sequence 10, Appl
6	703.5	15.0	798	1	US-08-200-900A-2	Sequence 2, Appli
7	703.5	15.0	798	5	PCT-US94-00616-2	Sequence 2, Appli
8	588	12.6	407	4	US-09-734-675-4	Sequence 4, Appli
9	560.5	12.0	492	4	US-09-685-166A-895	Sequence 895, App
10	558.5	11.9	492	3	US-09-342-749-2	Sequence 2, Appli
11	558.5	11.9	492	4	US-09-691-840-2	Sequence 2, Appli
12	547.5	11.7	235	3	US-08-944-483-65	Sequence 65, Appl
13	544.5	11.6	235	3	US-08-807-151-3	Sequence 3, Appli
14	544.5	11.6	235	4	US-09-478-957-3	Sequence 3, Appli

RESULT 1
US-09-027-337-2
; Sequence 2, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed
in
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides
; OTHER INFORMATION: 23 to 2589 of Sequence 1
; Patent No. 5972616
US-09-027-337-2

Query Match 100.0%; Score 4681; DB 2; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60

Qy         61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120

Qy        121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180

Qy        181 LPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240

Qy        241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQLCGTYPPS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQLCGTYPPS 300

Qy        301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGRLRKAQGTFNSPYYPGHY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGRLRKAQGTFNSPYYPGHY 360

Qy        361 PPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVTS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVTS 420

Qy        421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTGRCIRKELRCDGWADCTDH 480
```



RESULT 2
US-09-644-600-2
; Sequence 2, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-2

Query Match 100.0%; Score 4681; DB 4; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60

Qy         61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120

Qy        121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180

Qy        181 LPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240

Qy        241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQLCGTYPPS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQLCGTYPPS 300

Qy        301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGRLRKAQGTFNSPYYPGHY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGRLRKAQGTFNSPYYPGHY 360

Qy        361 PPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVTS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVTS 420

Qy        421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTGRCIRKELRCDGWADCTDH 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTGRCIRKELRCDGWADCTDH 480

Qy        481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSK 540

Qy        541 SQQCNGKDDCGDGSDEASCPKVNVVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SQQCNGKDDCGDGSDEASCPKVNVVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEK 600

Qy        601 DCDCGLRSFTRQARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 DCDCGLRSFTRQARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

Qy        661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720

Qy        721 AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLL 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLL 780

Qy        781 PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYT 840
```

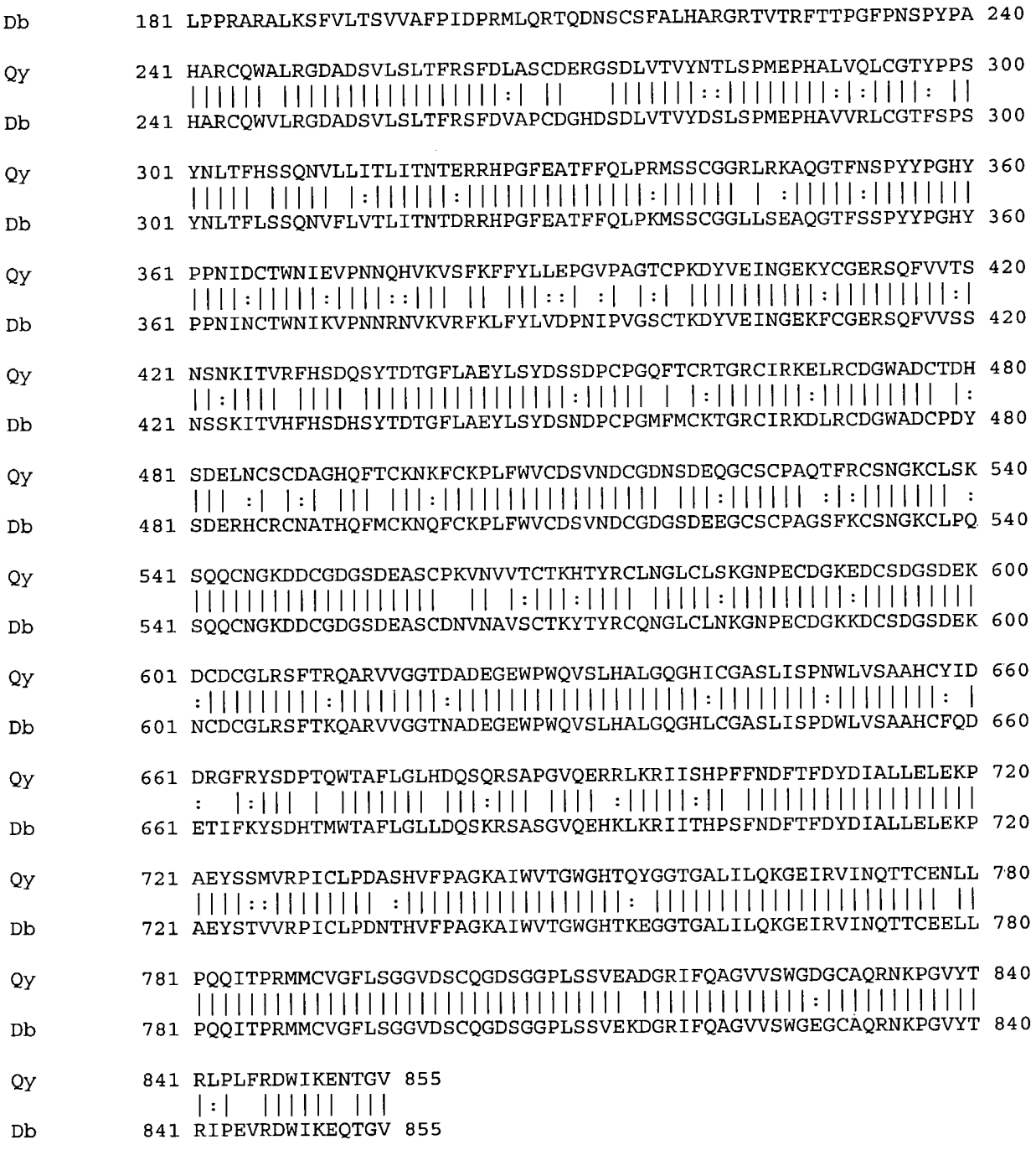
```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYT 840

Qy         841 RLPLFRDWIKENTGV 855
               ||||||||||||||||
Db         841 RLPLFRDWIKENTGV 855


RESULT 3
US-09-654-600A-2
; Sequence 2, Application US/09654600A
; Patent No. 6649741
; GENERAL INFORMATION:
;  APPLICANT: O'Brien, Timothy J.
;  APPLICANT:  Tanimoto, Hirotoshi
;  TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
;  TITLE OF INVENTION:  Overexpressed in Carcinomas
;  FILE REFERENCE: D6064CIP/D
;  CURRENT APPLICATION NUMBER: US/09/654,600A
;  CURRENT FILING DATE:  2000-09-01
;  PRIOR APPLICATION NUMBER: 09/421,213
;    09/027,337
;  PRIOR FILING DATE: 1999-10-20
;    1998-02-20
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
;   LENGTH: 855
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: TADG-15
US-09-654-600A-2

  Query Match             100.0%;  Score 4681;  DB 4;  Length 855;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  855;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60

Qy          61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120

Qy         121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180

Qy         181 LPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 LPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240

Qy         241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQLCGTYPPS 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQLCGTYPPS 300
```

```
Qy        301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGRLRKAQGTFNSPYYPGHY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGRLRKAQGTFNSPYYPGHY 360

Qy        361 PPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVTS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVTS 420

Qy        421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTGRCIRKELRCDGWADCTDH 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTGRCIRKELRCDGWADCTDH 480

Qy        481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSK 540

Qy        541 SQQCNGKDDCGDGSDEASCPKVNVVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SQQCNGKDDCGDGSDEASCPKVNVVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEK 600

Qy        601 DCDCGLRSFTRQARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 DCDCGLRSFTRQARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

Qy        661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720

Qy        721 AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLL 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLL 780

Qy        781 PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYT 840

Qy        841 RLPLFRDWIKENTGV 855
              |||||||||||||||
Db        841 RLPLFRDWIKENTGV 855
```

RESULT 1
JC7731
membrane-bound arginine-specific serine proteinase precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 03-Feb-2003
C;Accession: JC7731; JC7775
R;Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.;
Athauda, S.B.P.; Inoue, H.; Takahashi, K.
J. Biochem. 130, 425-430, 2001
A;Title: Characterization of a membrane-bound arginine-specific serine protease
from rat intestinal mucosa.
A;Reference number: JC7731; MUID:21421307; PMID:11530019
A;Accession: JC7731
A;Molecule type: mRNA
A;Residues: 1-855 <KIS>
A;Cross-references: DDBJ:AB049189
A;Experimental source: strain Male, 7-week-old
R;Satomi, S.; Yamasaki, Y.; Tsuzuki, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.
Biochem. Biophys. Res. Commun. 287, 995-1002, 2001
A;Title: A role for membrane-type serine protease (MT-SP1) in intestinal
epithelial turnover.
A;Reference number: JC7775; PMID:11573963
A;Contents: Small intestine
A;Accession: JC7775
A;Molecule type: mRNA
A;Residues: 1-855 <SAT>
A;Cross-references: DDBJ:AB037898
C;Comment: This enzyme, an epithelial-derived, type II integral membrane serine
protease. It localized mainly on brushborder membranes of the intestine and
participates in the processing or digestion of specific proteins or peptides on
the brushborder membranes. It also participates in the control of intestinal
epithelial turnover by regulating the cell-substratum adhesion associated with
epithelial migration and/or cell loss.
C;Genetics:
A;Gene: mt-sp1
A;Map position: basolateral cell surface
C;Superfamily: membrane-bound arginine-specific serine proteinase
C;Keywords: protein digestion

Query Match 83.0%; Score 3883; DB 2; Length 855;
Best Local Similarity 81.1%; Pred. No. 4.5e-247;
Matches 693; Conservative 79; Mismatches 83; Indels 0; Gaps 0;

```
Qy          1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
              ||::| || ||| :||||||||||| | :|| |||||||||||| |:||| || ||||: |
Db          1 MGNNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPVNNAKQVEKRGPRRWVVMVA 60

Qy         61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
              |:   ||: |  | |||| ||:||:||||||::||||||:|||||| ||||:||||:|
Db         61 VVFSFLLLSLMAGLLVWHFHYRNVRIQKVFNGHLRITNENFLDAYENSTSTEFISLASQV 120

Qy        121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180
              |:||||:|| || ||||||:| |||||||||||||||||| || || :| || ||||
Db        121 KEALKLMYSEVPVLGPYHKKSTVTAFSEGSVIAYYWSEFSIPPHLEEEVDRAMAVERVVT 180

Qy        181 LPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
              ||||||:|||||:||||||| | : :|||||||||| ||||| : ||||||||:|||||
```


Db 181 LPPRARALKSFVLTSVVAFPIDPRMLQRTQDNSCSFALHARGRTVTRFTTPGFPNSPYPA 240
Qy 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQLCGTYPPS 300
Db 241 HARCQWVLRGDADSVLSLTFRSFDVAPCDGHDSDLVTVYDSLSPMEPHAVVRLCGTFSPS 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGRLRKAQGTFNSPYYPGHY 360
Db 301 YNLTFLSSQNVFLVTLITNTDRRHPGFEATFFQLPKMSSCGGLLSEAQGTFSSPYYPGHY 360
Qy 361 PPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVTS 420
Db 361 PPNINCTWNIKVPNNRNVKVRFKLFYLVDPNIPVGSCTKDYVEINGEKFCGERSQFVVSS 420
Qy 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTGRCIRKELRCDGWADCTDH 480
Db 421 NSSKITVHFHSDHSYTDTGFLAEYLSYDSNDPCPGMFMCKTGRCIRKDLRCDGWADCPDY 480
Qy 481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSK 540
Db 481 SDERHCRCNATHQFMCKNQFCKPLFWVCDSVNDCGDGSDEEGCSCPAGSFKCSNGKCLPQ 540
Qy 541 SQQCNGKDDCGDGSDEASCPKVNVVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEK 600
Db 541 SQQCNGKDDCGDGSDEASCDNVNAVSCTKYTYRCQNGLCLNKGNPECDGKKDCSDGSDEK 600
Qy 601 DCDCGLRSFTRQARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 NCDCGLRSFTKQARVVGGTNADEGEWPWQVSLHALGQGHLCGASLISPDWLVSAAHCFQD 660
Qy 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720
Db 661 ETIFKYSDHTMWTAFLGLLDQSKRSASGVQEHKLKRIITHPSFNDFTFDYDIALLELEKP 720
Qy 721 AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLL 780
Db 721 AEYSTVVRPICLPDNTHVFPAGKAIWVTGWGHTKEGGTGALILQKGEIRVINQTTCEELL 780
Qy 781 PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYT 840
Db 781 PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEKDGRIFQAGVVSWGEGCAQRNKPGVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RIPEVRDWIKEQTGV 855

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3883	83.0	855	11	Q9JJI7	Q9jji7 rattus norv
2	2664	56.9	845	13	Q9DGR1	Q9dgr1 xenopus lae
3	2379	50.8	422	4	Q8WVC1	Q8wvc1 homo sapien
4	1011.5	21.6	572	11	Q8BIK6	Q8bik6 mus musculu
5	717.5	15.3	855	4	Q7Z410	Q7z410 homo sapien
6	717.5	15.3	1059	4	Q7Z411	Q7z411 homo sapien
7	690.5	14.8	1111	11	Q80YN4	Q80yn4 rattus norv
8	687	14.7	777	11	Q8CAN9	Q8can9 mus musculu
9	644.5	13.8	767	13	Q9DGR2	Q9dgr2 xenopus lae
10	636.5	13.6	680	5	Q868H7	Q868h7 branchiosto
11	623.5	13.3	680	5	Q868H5	Q868h5 branchiosto
12	617	13.2	581	5	Q9XZM7	Q9xzm7 strongyloce
13	612	13.1	688	5	Q868H6	Q868h6 branchiosto
14	601	12.8	490	11	Q7TN04	Q7tn04 mus musculu
15	600	12.8	490	11	Q920K3	Q920k3 rattus norv

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4681	100.0	855	10	US-09-776-191-2	Sequence 2, Appli
2	4681	100.0	855	12	US-10-072-012-352	Sequence 352, App
3	4681	100.0	855	12	US-10-072-012-411	Sequence 411, App
4	4681	100.0	855	12	US-10-072-012-418	Sequence 418, App
5	4681	100.0	855	14	US-10-099-700A-2	Sequence 2, Appli
6	4681	100.0	855	14	US-10-190-030B-2	Sequence 2, Appli
7	4681	100.0	855	14	US-10-302-840A-2	Sequence 2, Appli
8	4681	100.0	855	14	US-10-267-219-2	Sequence 2, Appli
9	4681	100.0	855	14	US-10-112-221A-2	Sequence 2, Appli
10	4681	100.0	855	14	US-10-104-271-2	Sequence 2, Appli
11	4681	100.0	855	15	US-10-147-211A-2	Sequence 2, Appli
12	4681	100.0	855	15	US-10-156-214A-2	Sequence 2, Appli
13	4681	100.0	855	16	US-10-600-187-2	Sequence 2, Appli
14	4676	99.9	855	12	US-10-072-012-353	Sequence 353, App
15	4676	99.9	855	12	US-10-072-012-412	Sequence 412, App
16	4676	99.9	855	12	US-10-072-012-419	Sequence 419, App
17	4676	99.9	855	15	US-10-295-027-1185	Sequence 1185, Ap
18	4672	99.8	855	12	US-10-072-012-354	Sequence 354, App
19	4672	99.8	855	12	US-10-072-012-420	Sequence 420, App
20	4672	99.8	855	12	US-10-037-417-132	Sequence 132, App
21	4631	98.9	851	12	US-10-276-774-1798	Sequence 1798, Ap
22	4631	98.9	851	12	US-10-296-115-1143	Sequence 1143, Ap
23	4175.5	89.2	782	14	US-10-097-340-312	Sequence 312, App
24	4175	89.2	762	16	US-10-729-807-1	Sequence 1, Appli
25	4111	87.8	757	12	US-10-072-012-44	Sequence 44, Appl
26	3901	83.3	855	9	US-09-900-751-2	Sequence 2, Appli
27	3901	83.3	855	12	US-10-072-012-355	Sequence 355, App
28	3901	83.3	855	12	US-10-072-012-413	Sequence 413, App

29	3883	83.0	855	12	US-10-072-012-356	Sequence 356, App
30	3883	83.0	855	12	US-10-072-012-414	Sequence 414, App
31	3883	83.0	855	12	US-10-072-012-417	Sequence 417, App
32	3810	81.4	902	12	US-10-333-743-3	Sequence 3, Appli
33	3810	81.4	902	16	US-10-600-187-10	Sequence 10, Appl
34	3810	81.4	902	16	US-10-297-987B-11	Sequence 11, Appl
35	2980	63.7	620	9	US-09-925-301-1193	Sequence 1193, Ap
36	2664	56.9	845	12	US-10-072-012-415	Sequence 415, App
37	1319	28.2	241	10	US-09-776-191-50	Sequence 50, Appl
38	1319	28.2	241	14	US-10-099-700A-4	Sequence

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4676	99.9	855	1	ST14_HUMAN	Q9y5y6 homo sapien
2	3901	83.3	855	1	ST14_MOUSE	P56677 mus musculu
3	1124.5	24.0	811	1	TMS6_MOUSE	Q9dbi0 mus musculu
4	1124	24.0	811	1	TMS6_HUMAN	Q8iu80 homo sapien
5	727	15.5	1034	1	ENTK_PIG	P98074 sus scrofa
6	712.5	15.2	1035	1	ENTK_BOVIN	P98072 bos taurus
7	692	14.8	1042	1	CORI_HUMAN	Q9y5q5 homo sapien
8	682.5	14.6	1019	1	ENTK_HUMAN	P98073 homo sapien
9	676.5	14.5	1069	1	ENTK_MOUSE	P97435 mus musculu
10	663.5	14.2	1113	1	CORI_MOUSE	Q9z319 mus musculu
11	600	12.8	490	1	TMS2_MOUSE	Q9jiq8 mus musculu
12	588	12.6	422	1	DES1_HUMAN	Q9ul52 homo sapien
13	586.5	12.5	704	1	CRAR_MOUSE	P98064 mus musculu
14	574	12.3	699	1	CRAR_HUMAN	P48740 h complemen
15	558.5	11.9	492	1	TMS2_HUMAN	O15393 homo sapien
16	546	11.7	453	1	TMS3_MOUSE	Q8k1t0 mus musculu
17	533.5	11.4	638	1	KAL_MOUSE	P26262 mus musculu
18	533	11.4	454	1	TMS3_HUMAN	P57727 homo sapien
19	518	11.1	603	1	CFAI_MOUSE	Q61129 mus musculu
20	518	11.1	604	1	CFAI_RAT	Q9wuw3 rattus norv
21	514.5	11.0	638	1	KAL_RAT	P14272 rattus norv
22	513	11.0	455	1	TMS5_MOUSE	Q9er04 mus musculu
23	511.5	10.9	418	1	HATT_HUMAN	O60235 homo sapien

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4681	100.0	855	2	AAY06671	Aay06671 Tumour an
2	4681	100.0	855	4	AAB98500	Aab98500 Human TAD
3	4681	100.0	855	4	AAE06930	Aae06930 Human mem
4	4681	100.0	855	5	AAO22929	Aao22929 Type II t
5	4681	100.0	855	6	ABP56619	Abp56619 Human mem
6	4681	100.0	855	6	AAO30146	Aao30146 Human mem
7	4681	100.0	855	6	AAE29820	Aae29820 Human mem
8	4681	100.0	855	6	AAE29791	Aae29791 Human mem
9	4681	100.0	855	6	ABP72376	Abp72376 Transmemb
10	4681	100.0	855	7	ADB97551	Adb97551 Human MTS
11	4676	99.9	855	3	AAB19552	Aab19552 Human mat
12	4676	99.9	855	4	AAB35465	Aab35465 Human mem
13	4631	98.9	851	4	AAM25628	Aam25628 Human pro
14	4631	98.9	851	4	ABB11428	Abb11428 Human mem
15	4319	92.3	932	4	ABG21442	Abg21442 Novel hum
16	4175.5	89.2	782	5	ABG96427	Abg96427 Human ova
17	4175	89.2	762	3	AAY90284	Aay90284 Human pep
18	3901	83.3	855	5	AAE23083	Aae23083 Epithin p
19	3810	81.4	902	4	AAB98507	Aab98507 Murine ep
20	3810	81.4	902	5	AAU80517	Aau80517 Mouse epi
21	3810	81.4	902	5	AAU77549	Aau77549 Murine ty
22	3781	80.8	683	3	AAB19551	Aab19551 Human mat
23	2980	63.7	620	3	AAB43748	Aab43748 Human can
24	1352	28.9	362	4	ABG21441	Abg21441 Novel hum

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3883	83.0	855	2	JC7731	membrane-bound arg
2	727	15.5	1034	1	A53663	enteropeptidase (E
3	712.5	15.2	1035	1	A43090	enteropeptidase (E
4	682.5	14.6	1019	1	A56318	enteropeptidase (E
5	663.5	14.2	1113	2	JE0315	low-density lipopr
6	578.5	12.4	1524	2	T30337	polyprotein - Afri
7	574	12.3	699	1	I54763	Ra-reactive factor
8	533.5	11.4	638	1	KQMSPL	plasma kallikrein
9	514.5	11.0	638	1	KQRTPL	plasma kallikrein
10	509.5	10.9	790	1	PLPG	plasmin (EC 3.4.21
11	506	10.8	613	2	S15468	complement C3b/C4b
12	502	10.7	460	2	B61545	plasmin (EC 3.4.21
13	501.5	10.7	786	1	A47547	serine proteinease
14	500	10.7	638	1	KQHUP	plasma kallikrein
15	497	10.6	810	1	PLHU	plasmin (EC 3.4.21
16	492.5	10.5	583	2	A29154	complement factor
17	491.5	10.5	812	1	PLMS	plasmin (EC 3.4.21
18	491	10.5	416	1	KFBO	coagulation factor
19	490.5	10.5	812	1	PLBO	plasmin (EC 3.4.21
20	490	10.5	417	1	S00845	hepsin (EC 3.4.21

```
RESULT 1
Q9JJI7
ID   Q9JJI7      PRELIMINARY;      PRT;   855 AA.
AC   Q9JJI7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Membrane bound serine protease (Membrane bound arginine specific
DE   serine protease).
GN   MBSP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=Jejunum;
RA   Tsuzuki S.;
RT   "A membrane bound serine protease expressed in rat small intestine.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar; TISSUE=Duodenum;
RA   Inoue H., Takahashi K., Kishi K.;
RT   "membrane-bound arginine specific serine protease.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DR   EMBL; AB037898; BAB03502.1; -.
DR   EMBL; AB049189; BAB13765.1; -.
DR   PIR; JC7731; JC7731.
DR   HSSP; P00763; 1DPO.
DR   MEROPS; S01.302; -.
DR   GO; GO:0004263; F:chymotrypsin activity; IEA.
DR   GO; GO:0008233; F:peptidase activity; IEA.
DR   GO; GO:0004295; F:trypsin activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR001254; Peptidase_S1.
DR   InterPro; IPR001314; Peptidase_S1A.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00192; LDLa; 3.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01209; LDLRA_1; 2.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Protease; Serine protease.
```

FT VARIANT 665 665 K -> N.
SQ SEQUENCE 855 AA; 94955 MW; 35806B7ECF6CF03D CRC64;

Query Match 83.0%; Score 3883; DB 11; Length 855;
Best Local Similarity 81.1%; Pred. No. 0;
Matches 693; Conservative 79; Mismatches 83; Indels 0; Gaps 0;

```
Qy          1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
              ||::| || ||| :||||||||||| | :|| |||||||||| |:||| || ||||: |
Db          1 MGNNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPVNNAKQVEKRGPRRWVVMVA 60

Qy         61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
              |:   ||: |  | |||| ||:||:||||||::|||||||:|||||| ||||:||||:|
Db         61 VVFSFLLLSLMAGLLVWHFHYRNVRIQKVFNGHLRITNENFLDAYENSTSTEFISLASQV 120

Qy        121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180
              |:||||:|| || ||||||:| |||||||||||||||||||| || || :| || ||||
Db        121 KEALKLMYSEVPVLGPYHKKSTVTAFSEGSVIAYYWSEFSIPPHLEEEVDRAMAVERVVT 180

Qy        181 LPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
              ||||||:|||||:||||||| | : :|||||||||| ||||| : ||||||||:|||||
Db        181 LPPRARALKSFVLTSVVAFPIDPRMLQRTQDNSCSFALHARGRTVTRFTTPGFPNSPYPA 240

Qy        241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQLCGTYPPS 300
              |||||| |||||||||||||||:| ||   |||||||::||||||||:|:||||: ||
Db        241 HARCQWVLRGDADSVLSLTFRSFDVAPCDGHDSDLVTVYDSLSPMEPHAVVRLCGTFSPS 300

Qy        301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGRLRKAQGTFNSPYYPGHY 360
              ||||| ||||| |:||||||:||||||||||||||:|||||| | :|||||:||||||||
Db        301 YNLTFLSSQNVFLVTLITNTDRRHPGFEATFFQLPKMSSCGGLLSEAQGTFSSPYYPGHY 360

Qy        361 PPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVTS 420
              ||||:|||||:||||::||| || |||::| :| |:| ||||||||||:|||||||||:|
Db        361 PPNINCTWNIKVPNNRNVKVRFKLFYLVDPNIPVGSCTKDYVEINGEKFCGERSQFVVSS 420

Qy        421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTGRCIRKELRCDGWADCTDH 480
              ||:|||| |||| ||||||||||||||:||||| | |:|||||||:|||||||| |:
Db        421 NSSKITVHFHSDHSYTDTGFLAEYLSYDSNDPCPGMFMCKTGRCIRKDLRCDGWADCPDY 480

Qy        481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSK 540
              ||| :| |:| ||| |||:|||||||||||||||| |||:|||||| :|:||||||| :
Db        481 SDERHCRCNATHQFMCKNQFCKPLFWVCDSVNDCGDGSDEEGCSCPAGSFKCSNGKCLPQ 540

Qy        541 SQQCNGKDDCGDGSDEASCPKVNVVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEK 600
              |||||||||||||||||| || |:|||:|||| |||||:|||||||||:|||||||||
Db        541 SQQCNGKDDCGDGSDEASCDNVNAVSCTKYTYRCQNGLCLNKGNPECDGKKDCSDGSDEK 600

Qy        601 DCDCGLRSFTRQARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
              :|||||||||:||||||||:|||||||||||||||||||:||||||||:||||||||: |
Db        601 NCDCGLRSFTKQARVVGGTNADEGEWPWQVSLHALGQGHLCGASLISPDWLVSAAHCFQD 660

Qy        661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720
              :  |:||| | ||||||| |||:||| |||| :|||||:|| ||||||||||||||||||
Db        661 ETIFKYSDHTMWTAFLGLLDQSKRSASGVQEHKLKRIITHPSFNDFTFDYDIALLELEKP 720

Qy        721 AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLL 780
              ||||::|||||||| :|||||||||||||||: |||||||||||||||||||||| ||
```

```
Db        721 AEYSTVVRPICLPDNTHVFPAGKAIWVTGWGHTKEGGTGALILQKGEIRVINQTTCEELL 780

Qy        781 PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYT 840
              ||||||||||||||||||||||||||||||| |||||||||||||:|||||||||||||
Db        781 PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEKDGRIFQAGVVSWGEGCAQRNKPGVYT 840

Qy        841 RLPLFRDWIKENTGV 855
              |:|  |||||| |||
Db        841 RIPEVRDWIKEQTGV 855
```

RESULT 2
ST14_MOUSE
ID   ST14_MOUSE   STANDARD;   PRT;   855 AA.
AC   P56677;
DT   15-JUL-1999 (Rel. 38, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Epithin).
GN   ST14 OR PRSS14.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C.B.17SCID; TISSUE=Thymus;
RX   MEDLINE=99216440; PubMed=10199918;
RA   Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,
RA   Schwartz R.H.;
RT   "Cloning and chromosomal mapping of a gene isolated from thymic
RT   stromal cells encoding a new mouse type II membrane serine protease,
RT   epithin, containing four LDL receptor modules and two CUB domains.";
RL   Immunogenetics 49:420-428(1999).
RN   [2]
RP   REVISIONS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.
RC   STRAIN=C.B.17SCID; TISSUE=Thymus;
RA   Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast tumor;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC   -!- TISSUE SPECIFICITY: Highly expressed in intestine, kidney, lung,
CC   and thymus. Not expressed in skeletal muscle, liver, heart,
CC   testis and brain.
CC   -!- SIMILARITY: Belongs to peptidase family S1.

```
CC   -!- SIMILARITY: Contains 2 CUB domains.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF042822; AAD02230.3; -.
DR   EMBL; BC005496; AAH05496.1; -.
DR   HSSP; P20231; 1AAO.
DR   MEROPS; S01.302; -.
DR   MGD; MGI:1338881; St14.
DR   GO; GO:0005576; C:extracellular; IDA.
DR   GO; GO:0019897; C:extrinsic to plasma membrane; IDA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IDA.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR001254; Peptidase_S1.
DR   InterPro; IPR001314; Peptidase_S1A.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01209; LDLRA_1; 2.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
KW   Transmembrane; Repeat.
FT   DOMAIN        1     55       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     56     76       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       77    855       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      214    331       CUB 1.
FT   DOMAIN      340    444       CUB 2.
FT   DOMAIN      451    488       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      489    522       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      523    561       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      565    604       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      615    854       SERINE PROTEASE.
FT   ACT_SITE    656    656       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    711    711       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    805    805       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    107    107       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    302    302       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    365    365       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    421    421       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    772    772       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   855 AA;  94654 MW;  4F10E84DA2146DD5 CRC64;

  Query Match             83.3%;  Score 3901;  DB 1;  Length 855;
  Best Local Similarity   81.8%;  Pred. No. 6.2e-261;
  Matches  699;  Conservative   73;  Mismatches   83;  Indels    0;  Gaps    0;

Qy          1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
              |||:| || ||| :|||||||||||| | :|| |||||||| || ||||| || ||||| |
Db          1 MGSNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPANNAKKVEKRGPRRWVVLVA 60

Qy         61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
              ||   ||: |  | ||||  ||:||||||||::|||||  |:||||||  ||||:||||:|
Db         61 VLFSFLLLSLMAGLLVWHFHYRNVRVQKVFNGHLRITNEIFLDAYENSTSTEFISLASQV 120

Qy        121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180
              |:|||||: || ||||||:|||||||||||||||||||||| || || :| || ||||
Db        121 KEALKLLYNEVPVLGPYHKKSAVTAFSEGSVIAYYWSEFSIPPHLAEEVDRAMAVERVVT 180

Qy        181 LPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
              ||||||:|||||:|||||||  | : :|||||||||| |||  |  : ||||||||:|||||
Db        181 LPPRARALKSFVLTSVVAFPIDPRMLQRTQDNSCSFALHAHGAAVTRFTTPGFPNSPYPA 240

Qy        241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQLCGTYPPS 300
              |||||| |||||||||||||||:| ||| ||||||||::||||||||:|:||||: ||
Db        241 HARCQWVLRGDADSVLSLTFRSFDVAPCDEHGSDLVTVYDSLSPMEPHAVVRLCGTFSPS 300

Qy        301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGRLRKAQGTFNSPYYPGHY 360
              ||||| ||||| |:|||||:|||||||||||||||:|||||| |   ||||:||||||||
Db        301 YNLTFLSSQNVFLVTLITNTDRRHPGFEATFFQLPKMSSCGGFLSDTQGTFSSPYYPGHY 360

Qy        361 PPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVTS 420
              ||||:|||||:||||::||| || |||::| || |:| |||||||||||||||||||:|
Db        361 PPNINCTWNIKVPNNRNVKVRFKLFYLVDPNVPVGSCTKDYVEINGEKYCGERSQFVVSS 420

Qy        421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTGRCIRKELRCDGWADCTDH 480
              ||:|||| |||| |||||||||||||||:||||| | |:|||||||||||||||| |:
Db        421 NSSKITVHFHSDHSYTDTGFLAEYLSYDSNDPCPGMFMCKTGRCIRKELRCDGWADCPDY 480

Qy        481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSK 540
              |||  | |:| |||||||:|||||||||||||||| |||:|||||| :|:||||||| :
Db        481 SDERYCRCNATHQFTCKNQFCKPLFWVCDSVNDCGDGSDEEGCSCPAGSFKCSNGKCLPQ 540

Qy        541 SQQCNGKDDCGDGSDEASCPKVNVVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEK 600
              ||:||||||:|||||||||  ||||:|||:|||| |||||||||||||| |||||||||
Db        541 SQKCNGKDNCGDGSDEASCDSVNVVSCTKYTYRCQNGLCLSKGNPECDGKTDCSDGSDEK 600

Qy        601 DCDCGLRSFTRQARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
              :|||||||||:||||||||:|||||||||||||||||||:||||||||:||||||||: |
Db        601 NCDCGLRSFTKQARVVGGTNADEGEWPWQVSLHALGQGHLCGASLISPDWLVSAAHCFQD 660

Qy        661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720
              |: |:||| | ||||||| |||:||| |||| :|||||:|| ||||||||||||||||
Db        661 DKNFKYSDYTMWTAFLGLLDQSKRSASGVQELKLKRIITHPSFNDFTFDYDIALLELEKS 720
```

```
Qy        721 AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLL 780
              |||::|||||||||:|||||||||||||||||: |||||||||||||||||||||:|:
Db        721 VEYSTVVRPICLPDATHVFPAGKAIWVTGWGHTKEGGTGALILQKGEIRVINQTTCEDLM 780

Qy        781 PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYT 840
              |||||||||||||||||||||||||||||| | |||:|||||||||:|||||||||||
Db        781 PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSAEKDGRMFQAGVVSWGEGCAQRNKPGVYT 840

Qy        841 RLPLFRDWIKENTGV 855
              |||: ||||||:|||
Db        841 RLPVVRDWIKEHTGV 855
```

```
RESULT 1
ST14_HUMAN
ID   ST14_HUMAN     STANDARD;      PRT;   855 AA.
AC   Q9Y5Y6; Q9BS01; Q9H3S0; Q9HB36; Q9HCA3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matriptase) (Membrane-
DE   type serine protease 1) (MT-SP1) (Prostamin) (Serine protease TADG-15)
DE   (Tumor associated differentially-expressed gene-15 protein).
GN   ST14 OR PRSS14 OR SNC19 OR TADG15.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99303581; PubMed=10373424;
RA   Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RT   "Molecular cloning of cDNA for matriptase, a matrix-degrading serine
RT   protease with trypsin-like activity.";
RL   J. Biol. Chem. 274:18231-18236(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99432178; PubMed=10500122;
RA   Takeuchi T., Shuman M.A., Craik C.S.;
RT   "Reverse biochemistry: Use of macromolecular protease inhibitors to
RT   dissect complex biological processes and identify a membrane-type
RT   serine protease in epithelial cancer and normal tissue.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RA   Yamaguchi N., Mitsui S.;
RT   "Molecular cloning of a novel transmembrane serine protease expressed
RT   in human prostate.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,
RA   O'Brien T.J.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood, and Muscle;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
```

RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 340-664 FROM N.A.
RA Cao J., Fan W., Zheng S.;
RT "Genomic analysis of a novel human serine protease SNC19.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN [7]
RP CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=99303582; PubMed=10373425;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matriptase
RT and a Kunitz-type serine protease inhibitor from human milk.";
RL J. Biol. Chem. 274:18237-18242(1999).
CC -!- FUNCTION: Degrades extracellular matrix. Proposed to play a role
CC in breast cancer invasion and metastasis. Exhibits trypsin-like
CC activity as defined by cleavage of synthetic substrates with Arg
CC or Lys as the P1 site.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC --

CC --
DR EMBL; AF118224; AAD42765.2; -.
DR EMBL; AF133086; AAF00109.1; -.
DR EMBL; AB030036; BAB20376.1; -.
DR EMBL; AF057145; AAG15395.1; -.
DR EMBL; BC005826; AAH05826.1; -.
DR EMBL; BC030532; AAH30532.1; -.
DR EMBL; AF283256; AAG13949.1; -.
DR HSSP; P00763; 1DPO.
DR Genew; HGNC:11344; ST14.
DR MIM; 606797; -.
DR MEROPS; S01.302; -.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0008236; F:serine-type peptidase activity; TAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.

```
DR   InterPro; IPR001314; Peptidase_S1A.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00192; LDLa; 3.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01209; LDLRA_1; 2.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
KW   Transmembrane; Repeat.
FT   DOMAIN        1     55       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     56     76       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       77    855       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      214    334       CUB 1.
FT   DOMAIN      340    447       CUB 2.
FT   DOMAIN      452    487       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      487    524       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      524    560       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      566    603       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      615    854       SERINE PROTEASE.
FT   ACT_SITE    656    656       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    711    711       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    805    805       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    109    109       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    302    302       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    485    485       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    772    772       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    327    329       FEA -> GTR (IN REF. 5; AAH05826).
FT   CONFLICT    381    381       R -> S (IN REF. 4).
FT   CONFLICT    674    674       A -> V (IN REF. 3).
SQ   SEQUENCE   855 AA;  94769 MW;  26143132C01F99C9 CRC64;

  Query Match             99.9%;  Score 4676;  DB 1;  Length 855;
  Best Local Similarity   99.9%;  Pred. No. 2.7e-314;
  Matches  854;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60

Qy         61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120

Qy        121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180

Qy        181 LPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
```